COMP1927 (16s1)

Ass2 (part-1) : How to Implement?

Notes:

- The document offers some **suggestions only**, with incomplete pseudo code
- The pseudo code is easy to read, but may not be efficient. You need to improve it!
- You can use code from labs/lecture material, however, must acknowledge it and provide a reference. For example you can use graph ADT implementation from one of the labs, and adapt it for this assignment.
- You can build a graph structure using Adjacency Matrix or List Representation.

readData.h

readData.c

(see next page)

graph.h

graph.c

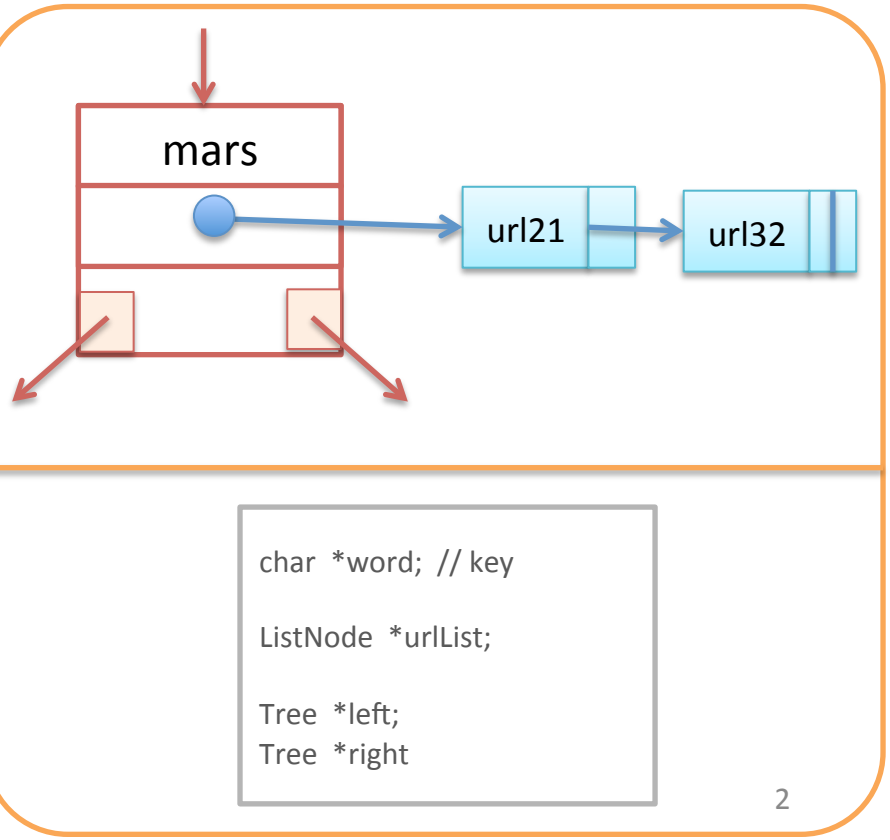
- See lectures and labs on Graph ADT

invertedIndex.h

invertedIndex.c

Possible implementations

- List of list
- BST where key is a string and value is a list (can use strcmp to compare key strings)



readData.c

List_of_Urls ← GetCollection( )

Create a set (list) of urls to process by reading data from file "collection.txt"

Graph g ← GetGraph(List_of_Urls)

Create empty graph (use graph ADT in say graph.h and graph.c)

For each url in the above list

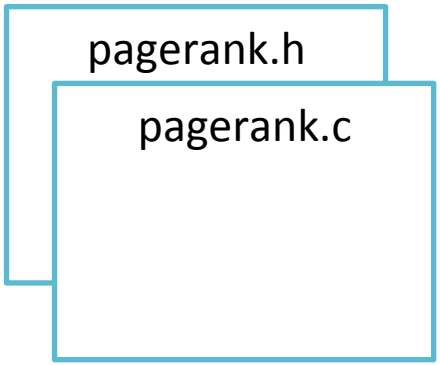
- read <url>.txt file, and update graph by adding a node and outgoing links

InvertedList ← GetInvertedList(List_of_Urls)

Create empty inverted list (use say List of lists, BST where values are lists, etc)

For each url in List_of_Urls

- read <url>.txt file, and update inverted index



pagerank.c

Get args : d, diffPR, maxIterations

List_of_Urls ← GetCollection( )
Graph g ← GetGraph(List_of_Urls)

List_Urls_PageRanks = calculatePageRank(g, d, diffPR, maxIterations);
Ordered_List_Urls_PageRanks = order (List_Urls_PageRanks)

Output Ordered_List_Urls_PageRanks to “pagerankList.txt”

invertedIndex.c

List_of_Urls ← GetCollection( )
InvertedIndex invertedIdx ← GetInvertedList (List_of_Urls)

Output invertedIdx to “invertedIndex.txt”

searchPagerank.h

searchPagerank.c

searchPagerank.c

Get query words from arguments

matched_Url_list ← findMatchedUrls(“invertedIndex.txt”, queryWords)
matched_Urls_with_PR ← findPagerank(“pagerankList.txt”, matched_Url_list)

Output ordered urls on stdout